International Journal of Agricultural Science and Research (IJASR)
ISSN(P): 2250-0057; ISSN(E): 2321-0087
Vol. 7, Issue 2, Apr 2017, 521-528

IDENTIFICATION OF RAINFALL PROBABILITY DISTRIBUTION FOR JUNAGADH

D. K. DWIVEDI[1], G. R. SHARMA[2] & D. V. PATEL[3]

[1]*Department of Soil & Water Engineering, College of Agricultural Engineering and Technology, Junagadh Agricultural University, Gujarat, India*

[2]*Polytechnic in Agricultural Engineering, Targhadia, Junagadh Agricultural University, Gujarat, India*

[3]*Department of Agricultural Statistics, College of Agriculture, Junagadh Agricultural University, Gujarat, India*

ABSTRACT

Rainfall data of 30 years (1986-2015) was analysed to obtain its descriptive statistics i.e. the mean, standard deviation, coefficient of variation, skewness, kurtosis, maximum value, minimum value and range. Four months of monsoon (June to September) and 17 weeks of the monsoon season were selected for identifying its probability distribution for which three tests namely, Anderson test, Kolmogorov-Smirnov and chi-square tests were carried out. By using the parameters of the selected distribution, random numbers were generated and the best distribution was identified based on minimum deviation between actual and estimated values. Gamma distribution was found to be the best fit distribution for the seasonal rainfall data while for most weeks generalized extreme value distribution was found to be the best fit distribution.

KEYWORDS: Probability Distribution, Rainfall, Goodness of Fit Tests, Kolmogorov Smirnov & Chi-Square

Received: Mar 18, 2017; **Accepted:** Apr 03, 2017; **Published:** Apr 10, 2017; **Paper Id.:** IJASRAPR201765

Original Article

INTRODUCTION

The rainfall distribution is important in designing soil conservation structures, water harvesting structures and watershed management strategies. The total rainfall received in each period at a location is highly variable from one year to another. The variability depends on the type of climate and the length of the considered period. Because of the strong variability of rainfall in time, the design and management of irrigation water supply and flood control systems are not based on the long-term average of rainfall records but on rainfall depths that can be expected for a specific probability. Although time series of rainfall data are characterized by their mean and standard variation, these values alone cannot beused to estimate design rainfall depths that can be expected with a specific probability. It is essential that the goodness of the assumed distribution be checked before carrying out further analysis of rainfall.

Phien and Ajiraja (1984) studied the application of log pearson type 3 distribution in hydrology and concluded that the log pearson type 3 distribution was applicable in most cases; however, for annual flood and maximum rainfall intensity, the existence of an upper bound to the distribution, in some cases, may cause some concerns, while this fact may indicate the suitability of the log pearson type 3 distribution for other variables such as annual stream flow and annual rainfall. Zalina*et al.* (2002) analysed the rainfall of Malaysia and concluded that the generalized extreme value distribution is the most appropriate distribution for describing the annual maximum rainfall series in Malaysia. Hanson and Vogel (2008) studied the probability distribution of daily rainfall in the

United States and the analysis indicated that the Pearson Type-III distribution fitted the full record of daily precipitation data remarkably well, while the Kappa distribution best describes the observed distribution of wet-day daily rainfall. Olofintoye*et al.* (2009) identified best fit probability distribution model for peak daily rainfall of selected cities in Nigeria. Their results showed that the log-Pearson type III distribution performed the best by occupying 50% of the total station number, while Pearson type III performed second best by occupying 40% of the total stations and lastly by log-Gumbel occupying 10% of the total stations. Sharma and Singh (2010) analysed the rainfall of Pantnagar using daily rainfall data set of 37 years. They found that lognormal and gamma distribution were the best fit probability distribution for the annual and monsoon season period of study, respectively and generalized extreme value distribution was observed in most of the weekly period as best fit probability distribution.

MATERIALS AND METHODS

The present study is based on rainfall data of 30 years (1986 to 2015) observed at Junagadh located in Gujarat State of India. Geographically Junagadh is situated at $21^0$31' N latitude and $70^0$28' E longitudes with an elevation of 107m above M.S.L. Four monsoon months (June, July, August and September) and 17 weeks were selected for analysis. Descriptive statistics the mean, standard deviation, coefficient of variation, skewness, kurtosis, maximum value, minimum value and range for the rainfall data were obtained using data analysis tool of excel.

The probability distributions viz. normal, lognormal (2P, 3P), gamma (2P, 3P), generalized gamma (3P, 4P), log-gamma, weibull (2P, 3P), pearson 5 (2P, 3P), pearson 6 (3P, 4P), log-pearson and generalized extreme value were selected to find the best fit probability distribution for rainfall. The goodness of fit test were used to measure the compatibility of random sample with the theoretical probability distribution.

The goodness of fit tests was applied for testing the following null hypothesis:

H_0: The rainfall data follows the specified distribution

H_a: The rainfall data does not follow the specified distribution.

The following goodness-of-fit tests viz. Kolmogorov-Smirnov test , Anderson-Darling test and the chi-square test at α (0.01) level of significance were used for the selection of the best fit Probability distribution.

Kolmogorov-Smirnov Test

In statistics, the Kolmogorov–Smirnov test is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution . The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The Kolmogorov-Smirnov statistic (D) is defined as the largest vertical difference between the theoretical and the empirical cumulative distribution function (ECDF):

$$D = \max_{1\leq i\leq n}\left(F(x_i)-\frac{i-1}{n},\frac{i}{n}-F(x_i)\right)$$

Where X_i = random sample, i =1, 2,….., n.

$$CDF = F_n(x) = \frac{1}{n}[Number\ of\ observations\ \leq x]$$

This test is used to decide if a sample comes from a hypothesized continuous distribution.

Anderson-Darling Test

The Anderson–Darling test is a statistical test of whether a given sample of data is drawn from a given probability distribution. In its basic form, the test assumes that there are no parameters to be estimated in the distribution being tested, in which case the test and its set of critical values is distribution-free. However, the test is most often used in contexts where a family of distributions is being tested, in which case the parameters of that family need to be estimated and account must be taken of this in adjusting either the test-statistic or its critical values The Anderson-Darling statistic (A^2) is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(x_i) + \ln(1 - F(x_{n-i+1}))]$$

It is a test to compare the fit of an observed cumulative distribution function to an expected cumulative distribution function. This test gives more weight to the tails then the Kolmogorov-Smirnov test.

Chi-Square Test

This test is for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution.

The Chi-Squared statistic is defined as

$$x^2 = \sum_{i=1}^{k}\frac{(O_i - E_i)^2}{E_i}$$

Where,

Oi = observed frequency,

E_i= expected frequency,

'i'= number of observations (1, 2, …….k)

E_i is calculated by the following computation

$E_i = F(x_2) - F(x_1)$

F is the CDF of the probability distribution being tested.

The observed number of observation (k) in interval 'i' is computed from equation given below

$K = 1 + \log_2 n$ where n is the sample size.

The three goodness of fit test was fitted to the rainfall data. The test statistic of each test was computed and tested at (α =0.01) level of significance. The ranking of different probability distributions will be marked from 1 to 16 based on minimum test statistic value. The distribution holding the first rank will be selected for all the three tests independently. The assessments of all the probability distribution will be made on the bases of total test score obtained by combining the entire three tests. The random numbers will be generated for the distributions and residuals (R) will be computed for each observation of the data set using least squares.

$$R = \left|\sum_{i-1}^{n}(Y - \hat{Y}_I)\right|^2$$

Where, Y = the actual observation

Y_I= the estimated observation (i = 1, 2,….., n)

The distribution having minimum sum of residuals was considered as the best fit probability distribution for that particular data set.

RESULTS AND DISCUSSIONS

Rainfall analysis of 30 years (1986-2015) was carried out as explained in the methodology above. The results of the descriptive statistics obtained is given in Table 1 and Table 2

Table 1: Summary of Descriptive Statistics for Weekly Rainfall

Study Period (1986-2015)		Mean (mm)	S.D (mm)	C.V (%)	Kurtosis	Skewness	Range (mm)	Minimum (mm)	Maximum (mm)
	Seasonal	56.8	82.9	1.46	6.4	2.3	444.2	0.0	444.2
June	1 week	6.9	15.9	2.31	5.4	2.5	60.9	0.0	60.9
June	2 week	7.3	15.5	2.11	7.8	2.7	67.2	0.0	67.2
June	3 week	63.6	96.5	1.52	6.1	2.3	426.0	0.0	426.0
June	4 week	54.0	90.3	1.67	6.4	2.4	398.0	0.0	398.0
July	5 week	37.9	68.8	1.81	4.3	2.3	246.4	0.0	246.4
July	6 week	62.3	70.5	1.13	0.0	1.1	234.2	0.0	234.2
July	7 week	71.9	82.3	1.15	1.7	1.5	304.1	0.0	304.1
July	8 week	86.7	106.8	1.23	3.2	1.8	410.9	1.1	412.0
August	9 week	89.3	92.8	1.04	1.4	1.3	359.4	0.0	359.4
August	10 week	77.0	99.0	1.28	-0.1	1.1	298.6	0.0	298.6
August	11 week	58.0	85.1	1.47	7.0	2.5	382.4	0.0	382.4
August	12 week	40.9	76.5	1.87	20.3	4.3	411.9	0.0	411.9
September	13 week	24.7	26.7	1.08	1.1	1.4	88.8	0.0	88.8
September	14 week	36.8	55.4	1.50	6.5	2.4	244.6	0.0	244.6
September	15 week	37.3	53.7	1.44	0.6	1.3	174.2	0.0	174.2
September	16 week	54.0	100.6	1.86	8.9	2.9	444.2	0.0	444.2
October	17 week	32.0	39.7	1.24	-0.4	1.0	119.8	0.0	119.8

S.D – Standard Deviation, C.V – Coefficient of Variation

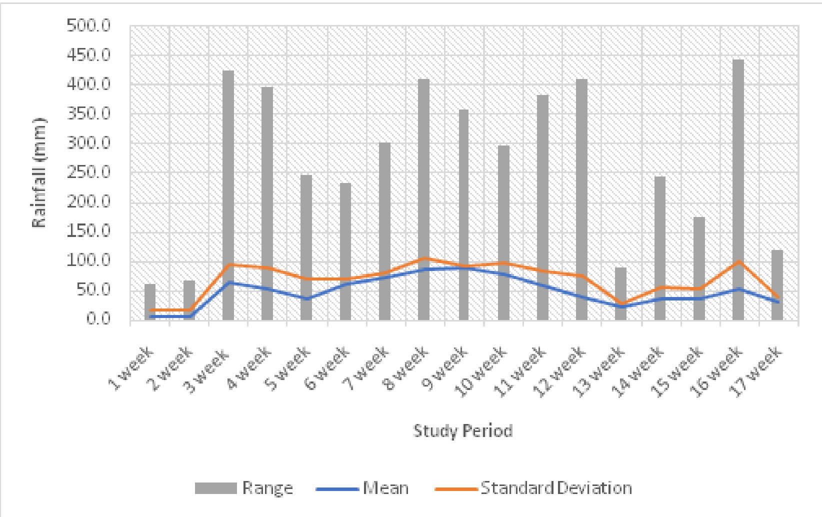

Figure 1: Mean, Standard Deviation and Range of Weekly Rainfall

The mean of seasonal weekly rainfall of 30 years was obtained as 56.80 mm. The maximum value of weekly rainfall mean was found to be 89.3 mm and it was obtained in the first week of August. The maximum amount of rainfall was found to be 444.2 in the last week of September. The standard deviation for weekly seasonal rainfall for 30 years was 82.9 mm while for weekly rainfall standard deviation ranged from 15.6 mm in second week of June to 106.8 mm in fourth week of July. The graphical representation of the weekly rainfall is shown in Figure 1.

Table 2: Summary of Descriptive Statistics of Monthly Seasonal Rainfall

Study Period (1986-2015)	Mean (mm)	S.D (mm)	C.V (%)	Kurtosis	Skewness	Range (mm)	Minimum (mm)	Maximum (mm)
June	173.18	142.1	0.82	1.006	1.089	521.4	15.7	537.1
July	340.88	233.76	0.69	1.6999	1.0285	1025.8	30.6	1056.4
August	200.86	170.64	0.85	0.8105	1.2046	675.7	3.7	679.4
September	160.97	141.74	0.88	0.9277	1.1364	525.1	7.8	532.9
Seasonal	875.89	357.521	0.41	0.61204	0.052827	1419.8	138.2	1558

S.D – Standard Deviation, C.V – Coefficient of Variation

The maximum mean of monthly rainfall was obtained as 340.88 in the month of July. The rainfall ranged from minimum of 3.7 mm to maximum value of 1056.4 mm. Maximum standard deviation of 233.76 mm was also found in the month of July. The mean of seasonal monthly rainfall was obtained as 875.89 mm with standard deviation of 357.521 mm. The range of seasonal rainfall varied from a minimum value of 138.2 mm to a maximum value of 1558 mm. The graphical representation of mean, standard deviation and range of monthly rainfall is shown in Figure 2.

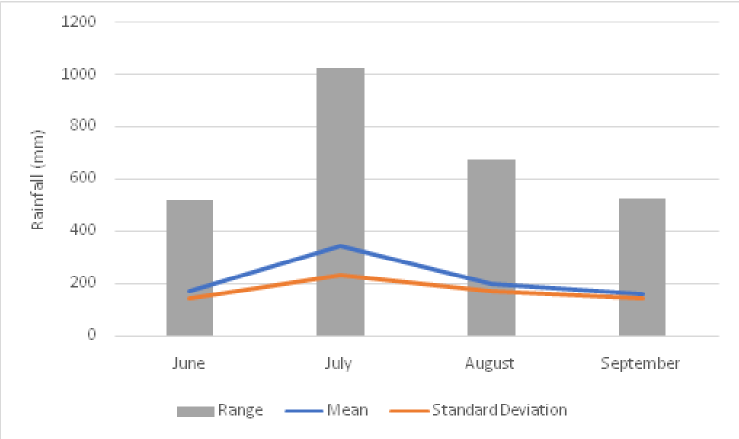

Figure 2: Mean, Standard Deviation and Range of Monthly Rainfall

Table 3: Distributions Fitted for Rainfall Data Set

	Kolmogorov Smirnov		Anderson Darling		Chi-Square	
Study Period	**Distribution**	**Statistic**	**Distribution**	**Statistic**	**Distribution**	**Statistic**
Seasonal	Gen. Extreme	0.1593	Gen. Extreme	18.695	Gamma	72.913
1 week	Normal	0.4341	Weibull	-1.4333	Normal	10.9
2 week	Gen. Extreme	0.3309	Weibull	-3.7627	Gen. Extreme	4.536
3 week	Gen. Extreme	0.1688	Gen. Extreme	1.2195	Gen. Extreme	1.7281
4 week	Gamma	0.1667	Gen. Extreme	1.5461	Gamaa	1.0326
5 week	Gen. Extreme	0.2043	Gen. Extreme	1.9753	Gamaa	1.5985
6 week	Gen. Extreme	0.1388	Gen. Extreme	0.9178	Gen. Extreme	0.5722
7 week	Lognormal	0.0975	Gen. Extreme	0.6669	Lognormal	1.0335
8 week	Gen. Gamma (4P)	0.0737	Gen. Gamma	0.2115	Gen. Gamma	0.2912
9 week	Lognormal	0.1084	Gen. Extreme	0.4934	Gen. Extreme	1.1024
10 week	Lognormal	0.1385	Gen. Extreme	1.8398	Gen. Extreme	1.8525
11 week	Lognormal	0.0914	Gen. Extreme	0.4819	Pearson 5	0.0934
12 week	Gen. Extreme	0.0776	Gen. Extreme	0.2776	Gen. Extreme	0.1609
13 week	Gamma	0.1313	Gen. Extreme	0.6917	Gen. Extreme	2.8382
14 week	Gen. Extreme	0.1335	Gen. Extreme	0.7347	Weibull	0.2665
15 week	Gen. Extreme	0.2948	Gen. Extreme	3.448	Gen. Extreme	6.0381
16 week	Gen. Extreme	0.1879	Gen. Extreme	1.3197	Gamma	0.4735
17 week	Gen. Extreme	0.2106	Gen. Extreme	1.7996	Normal	2.8414
June	Gen. Extreme	0.1085	Gen. Extreme	0.3898	Gen. Extreme	0.0317

Table 3: Contd.,						
July	Gen. Extreme	0.0995	Gen. Extreme	0.3311	Pearson 5 (3P)	0.0813
August	Lognormal (3P)	0.0792	Lognormal (3P)	0.1785	Pearson 6 (4P)	0.1639
September	Gen. Extreme	0.0923	Gen. Extreme	0.3149	Gen. Gamma	0.2019

Table 4: Parameters of the Distributions Fitted for Rainfall Data Sets

Study Period	Distributions	Parameters
Seasonal	Generalized Ext. Value	α=0.60537 β=103.1
	Gamma	k=0.47748 σ=24.854 μ=13.118
Week 1	Normal	σ=15.941 μ=6.9
	Weibull	α=0.2773 β=1.0085
Week 2	Generalized Ext. Value	k=0.70204 σ=2.3016 μ=0.74685
	Weibull	α=0.27686 β=1.0986
Week 3	Generalized Ext. Value	k=0.47729 σ=32.369 μ=16.3
Week 4	Generalized Ext. Value	k=0.54453 σ=24.687 μ=11.223
	Gamma	α=0.45811 β=141.56
Week 5	Generalized Ext. Value	k=0.60775 σ=15.283 μ=6.1625
	Gamma	α=0.4533 β=109.13
Week 6	Generalized Ext. Value	k=0.25691 σ=40.437 μ=25.325
Week 7	Lognormal	σ=1.4607 μ=3.5954
	Generalized Ext. Value	k=0.30388 σ=42.571 μ=29.245
Week 8	Generalized Gamma (4P)	k=1.0188 α=0.53646 β=161.48 γ=1.1
	Generalized Gamma	k=0.98198 α=0.66417 β=131.71
Week 9	Lognormal	σ=1.5543 μ=3.8605
	Generalized Ext. Value	k=0.22035 σ=55.7 μ=41.771
Week 10	Lognormal	σ=1.9654 μ=3.2401
	Generalized Ext. Value	k=0.34683 σ=47.654 μ=24.988
Week 11	Lognormal	σ=1.4215 μ=3.2141
	Generalized Ext. Value	k=0.50721 σ=25.71 μ=17.558
	Pearson 5	α=0.64361 β=6.1918
Week 12	Generalized Ext. Value	k=0.57089 σ=15.681 μ=11.623
Week 13	Gamma	α=0.85311 β=28.914
	Generalized Ext. Value	k=0.24852 σ=15.106 μ=11.084
Week 14	Weibull	α=0.21734 β=15.871
	Generalized Ext. Value	k=0.50419 σ=17.041 μ=10.198
Week 15	Generalized Ext. Value	k=0.41828 σ=21.847 μ=9.5045
Week 16	Generalized Ext. Value	k=0.60352 σ=21.475 μ=9.8985
	Gamma	α=0.47824 β=141.05
Week 17	Normal	σ=39.707 μ=31.98
	Generalized Ext. Value	k=0.2972 σ=21.174 μ=11.06
June	Generalized Ext. Value	k=0.0881 σ=103.29 μ=103.76
July	Generalized Ext. Value	k=0.00948 σ=187.1.29 μ=234.62
	Pearson 5 (3P)	α=0.60537 β=103.1 γ = 454.68
August	Lognormal (3P)	σ=0.7479 μ=5.174 γ = -29.25
	Pearson 6 (4P)	α_1=18.471 α_2=3.5561 β=37.33 γ=-63.233

Table 4: Contd.,		
September	Generalized Ext. Value	k=0.13336 σ=97.561 μ=89.952
	Generalized Gamma	k=0.93523 α=1.2233 β=124.82

Table 5: Best-Fit Probability Distribution for Rainfall

Study Period	Best-Fit
Seasonal	Gamma
1 week	Normal
2 week	Gen. Extreme Value
3 week	Gen. Extreme Value
4 week	Gamma
5 week	Gamma
6 week	Gen. Extreme Value
7 week	Lognormal
8 week	Generalized Gamma
9 week	Lognormal
10 week	Lognormal
11 week	Lognormal
12 week	Gen. Extreme Value
13 week	Gamma
14 week	Gen. Extreme Value
15 week	Gen. Extreme Value
16 week	Gamma
17 week	Normal
June	Gen. Extreme Value
July	Pearson 5 (3P)
August	Pearson 6 (4P)
September	Gen. Gamma

The weekly rainfall data was analysed by testing all the 16 probability distributions for best fit using the three goodness of fit tests mentioned above. The probability distributions were grouped according to the rank obtained in all the three tests and three probability distribution with first rank obtained from the three tests were selected. The random numbers were generated for actual and estimated observations using the parameters of the probability distribution shown in Table 5 and the residuals were computed for each data set. Sum of these deviation were obtained for all the identified probability distribution. The probability distribution having minimum deviation was treated as the best selected probability distribution for the individual data set. The best fit probability distribution for each of the selected week is given Table 5. Generalized extreme value distribution was found common in most weeks followed by gamma distribution. Pearson 5 (3P) and Pearson 6 (4P) distribution was obtained as best fit distribution for July and August respectively while rainfall in September followed generalized gamma distribution.

CONCLUSIONS

The 30-year (1986-2015) rainfall data of Junagadh was analysed to obtain the descriptive statistics. The mean, standard deviation, skewness, kurtosis, range, maximum value and minimum value of weekly data and monthly data of June, July, September and August were obtained. The maximum amount of rainfall was found to be 444.2 in the last week of September. The maximum mean of monthly rainfall was obtained as 340.88 in the month of July. The range of seasonal rainfall varied from a minimum value of 138.2 mm to a maximum value of 1558 mm. The three tests namely Anderson, Smirnov and Chi-square can be reliably used to obtain the probability distributions. By using the parameters of the selected distribution, random numbers can be generated and the best distribution can be identified based on minimum deviation

between actual and estimated values. The analysis for probability distribution revealed that generalized extreme value distribution was the most common probability distribution of the weekly rainfall data.

REFERENCES

1. *Daniel, Wayne W. (1990)."Kolmogorov–Smirnov one-sample test". Applied Nonparametric Statistics (2nd ed.). Boston: PWS-Kent. pp. 319–330.*
2. *Greenwood, P.E. and Nikulin, M.S. (1996) A guide to chi-squared testing. Wiley, New York. ISBN 0-471-55779-X*
3. *Hanson, L. S., & Vogel, R. (2008). The probability distribution of daily rainfall in the United States. In World Environmental and Water Resources Congress 2008: Ahupua'A (pp. 1-10).*
4. *Olofintoye, O. O., Sule, B. F., & Salami, A. W. (2009). Best-fit probability distribution model for peak daily rainfall of selected cities in Nigeria. New York Science Journal, 2(3), 1-12.*
5. *Phien, H. N., &Ajirajah, T. J. (1984). Applications of the log Pearson type-3 distribution in hydrology. Journal of hydrology, 73(3-4), 359-372.*
6. *Sharma, M. A., & Singh, J. B. (2010). Use of probability distribution in rainfall analysis. New York Science Journal, 3(9), 40-49.*
7. *Stephens, M. A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons". Journal of the American Statistical Association.* ***69****: 730–737.*
8. *Zalina, M. D., Desa, M. N. M., Nguyen, V. T. A., &Kassim, A. H. M. (2002). Selecting a probability distribution for extreme rainfall series in Malaysia. Water science and technology, 45(2), 63-68.*